



Contents lists available at ScienceDirect

## Journal of Experimental Child Psychology

journal homepage: [www.elsevier.com/locate/jecp](http://www.elsevier.com/locate/jecp)



# Lumping and splitting: Developmental changes in the structure of children's semantic networks

Catarina Vales\*, Patience Stevens, Anna V. Fisher

Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, USA



### ARTICLE INFO

#### Article history:

Received 9 December 2019

Revised 6 April 2020

Accepted 24 May 2020

Available online 22 July 2020

#### Keywords:

Semantic structure

Semantic differentiation

Knowledge

Learning

Cognitive development

### ABSTRACT

Organized semantic representations encoding across- and within-domain distinctions are a hallmark of mature cognition, and understanding how they change with experience and learning is a key endeavor in developmental science. Existing computational modeling studies provide a mechanistic framework for understanding how structured semantic representations emerge as a result of development and learning. However, their predictions remain largely untested in young children, with the existing evidence providing only indirect tests of these predictions. Across two experiments, we provide the first direct examination of a key prediction derived from these computational models—that early in development, broad across-domain distinctions should generally be more strongly represented relative to finer-grained within-domain distinctions. The results support this hypothesis, being consistent with the exploitation of patterns of covariation among entities as a mechanism supporting the acquisition of structured semantic representations.

© 2020 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Introduction

Understanding a story one is reading, following the plot of a movie, and making predictions about entities that surround us—these, and many other cognitive processes, rely on semantic knowledge, namely one's knowledge of word meanings, object features, and facts (Clark, 1973). There is broad

\* Corresponding author.

E-mail address: [cvalues@andrew.cmu.edu](mailto:cvalues@andrew.cmu.edu) (C. Vales).

agreement that semantic knowledge is more than a collection of isolated representations of word and object meanings. Rather, contemporary theoretical accounts agree that our knowledge is structured in *semantic networks* that encode concepts according to relevant within- and across-domain distinctions (Cree & McRae, 2003; Crowe & Prescott, 2003; McClelland & Rogers, 2003; McRae & Jones, 2013). There is much evidence that these semantic networks play an important role in supporting a broad range of cognitive processes (Bjorklund & Jacobs, 1985; Borovsky, Ellis, Evans, & Elman, 2016; Colunga & Sims, 2017; Federmeier & Kutas, 1999; Fisher, Godwin, & Matlen, 2015; Gobbo & Chi, 1986; Kaduk et al., 2016; Medin, Lynch, Coley, & Atran, 1997), making the development of human semantic representations a central question in the study of cognition.

A number of computational modeling studies provide a comprehensive framework for understanding how semantic differentiation emerges as a result of development and learning. However, many of their key predictions remain largely untested in young children, with the existing studies providing only indirect evidence. In what follows, we first present a framework derived from those computational modeling studies and then review the existing empirical evidence addressing one of their key predictions. We then report two experiments designed to address the limitations of prior empirical studies to directly evaluate that prediction.

### *The development of semantic structure: A mechanistic framework derived from computational modeling studies*

The properties of the entities in the world are not arbitrarily clustered; instead, there are higher-order relations among their properties that can in principle be capitalized on to learn relevant within- and across-domain distinctions (Malt & Smith, 1984; McRae, Cree, Westmacott, & De Sa, 1999; McRae, De Sa, & Seidenberg, 1997; Rosch & Mervis, 1975). For example, entities that can move tend to also breathe, grow, and reproduce, and entities that have feathers tend to also have beaks and wings, build nests, and have the ability to fly. A literature that focuses on naturally occurring group differences, both in adults and in children, suggests that these clusters of features are likely learned from experience with items in a domain. For example, fishermen sort pictures of fish into a larger number of meaningful groups relative to novices (Boster & Johnson, 1989), and the focus of one's expertise in a domain (e.g., botanists vs. landscape workers in the domain of plants) determines one's semantic structure in that domain (Medin et al., 1997). Analogous results are seen in children who differ in their early experiences with specific domains of knowledge such as children who show a special interest in a domain (e.g., dinosaurs) relative to children who do not (Gobbo & Chi, 1986), children who grow up in a rural setting relative to children who grow up in an urban setting (Coley, 2012), and children who own a pet relative to children who do not (Inagaki, 1990).

This idea—that experience-based learning of clusters of features leads to the emergence of semantic differentiation—finds support across a number of computational modeling studies, which together provide a mechanistic framework for understanding how semantic representations differentiate as a result of experience (Hills, Maouene, Maouene, Sheya, & Smith, 2009; Kemp & Tenenbaum, 2008; McClelland & Rogers, 2003). For example, McClelland and Rogers (2003; see also Rogers & McClelland, 2004) showed that the internal representations of a neural network became gradually more differentiated—first into broad domains such as plants and animals and then into more fine-grained within-domain distinctions such as flowers and trees—as the network learned about the features of the items in the training set (e.g., can move, can fly). Using a graph-theoretic modeling approach, Hills et al. (2009) also showed that the features normatively associated with the nouns that most children know by 3 years of age provide sufficient information to aggregate those words into clusters that resemble adult-like domains (e.g., food, animals, vehicles, clothes) (see Peters & Borovsky, 2019, and Sizemore, Karuza, Giusti, & Bassett, 2018, for related evidence). Importantly, as the number of available features increased—presumably as happens with increased experience—so did the differentiation among the items into identifiable clusters. A similar increase in differentiation among items within the domain of animals with increasingly available features was also reported by Kemp and Tenenbaum (2008), who used a hierarchical Bayesian approach to model the emergence of structured knowledge.

Although these models are computationally distinct, their results converge with the idea that the gradual accumulation of experiences with items of a domain, as well as their properties, supports the emergence of across- and within-domain differentiation. The models—and, by hypothesis, human children—acquire differentiated representations by capitalizing on the clusters of features that differentiate items across domains and within a domain (e.g., “can move” applies to animals but not to plants, only some animals “can fly”). This reliance on the coherent covariation among entity features makes a key prediction regarding semantic differentiation early in development—specifically, that broad across-domain distinctions should generally be acquired earlier and thus be more strongly represented relative to finer-grained within-domain distinctions. This is because entities belonging to different domains tend to differ on largely nonoverlapping clusters of features, whereas entities within a domain tend to differ on partially overlapping features. For example, consider the following items: owl, salmon, oak, and daisy. The distinction between the domains of animals (owl and salmon) and plants (oak and daisy) relies on a number of features that coherently covary in tandem with little overlap—both animals have eyes and can move, and both plants have leaves and roots, and there is not much feature overlap across the two domains. On the other hand, differentiating within each of those domains relies on partially overlapping features, which still covary with the features characterizing most items from that domain. That is, although both owl and salmon have eyes and can move, differentiating within this domain requires learning that one animal has feathers and builds nests and the other one has scales and lives in water—but each of these features still consistently covaries with the presence of eyes and mobility. Similarly, although both plants have leaves and roots, differentiating between a tree and a flower requires learning that one has branches and the other has petals—and these features still covary with the presence of leaves and roots at the item level. In other words, properties that discriminate items within a domain (e.g., scales, feathers) tend to covary with properties that are shared across all items of that domain (e.g., eyes, mobility), making the acquisition of within-domain distinctions more challenging and thus generally emerging more slowly relative to across-domain distinctions. In sum, this framework predicts that, early in development, differentiation within domains of knowledge should generally be weaker relative to across-domain differentiation. We now turn to the existing empirical evidence that speaks to this prediction and the open questions that remain.

#### *Empirical studies of semantic differentiation: Existing evidence and open questions*

The existing empirical studies provide only *indirect* evidence regarding the prediction that, early in development, within-domain differentiation should be weaker relative to across-domain differentiation. For example, studies using the sequence in which infants and toddlers inspect sets of items suggest that although their semantic networks are broadly differentiated into meaningful domains such as animals, vehicles, and foods (Mandler & Bauer, 1988; Mandler, Bauer, & McDonough, 1991; Mandler & McDonough, 1993; Pauen, 2002), additional experience might be required for the emergence of within-domain differentiation. Consistent with this last point, studies using a property attribution procedure suggest that whereas kindergartners are equally likely to accept a property as being true of items of the same domain (e.g., that both a man and a pig can “be sorry”), sixth graders show a more graded pattern of attributions (e.g., that a man, but not a pig, can “be sorry”) (Keil, 1979; see also Carey, 1985)—a behavior consistent with increased within-domain differentiation over development.

A number of studies have also used a match-to-sample procedure—in which children are asked to indicate which of a few options matches or shares an unobservable property with a target item—to investigate the kinds of relations among items that children of different ages prioritize (e.g., Smiley & Brown, 1979; Walsh, Richardson, & Faulkner, 1993; Waxman & Namy, 1997). This literature has provided considerable evidence for a developmental shift in children’s preferences from context-based items (e.g., selecting *rabbit* when the target is *carrot*) to same-domain items (e.g., selecting *tomato* for the same *carrot* target). Although these results do not directly speak to the acquisition of across- and within-domain distinctions, they suggest that same-domain items might not be as strongly linked early in development as the framework discussed above proposes. However, there is also evidence that children’s momentary preferences in this task are modulated by in-task contextual factors. For example, even young children can show equal preferences for context-based and same-domain

matches if, instead of needing to choose between the two, they are asked to choose between one of those and an unrelated item (Nguyen, 2007; Nguyen & Murphy, 2003); similarly, the wording used to prompt children to make a selection has also been shown to influence their choices (Waxman & Namy, 1997).

Taken together, the existing evidence for the emergence of across- and within-domain differentiation in children's semantic structure presents three main limitations. First, most of these tasks generate binary semantic judgments (e.g., by forcing children to select one item or to decide whether to extend a predicate) among a small number of items, but semantic representations differentiate many items along multiple relations varying in strength. Second, by using different tasks across different ages, the age-related changes observed in these studies may have been, at least in part, confounded with task-specific factors. Finally, these tasks likely provide only indirect measures of semantic structure. For example, in addition to the effects of contextual factors in the match-to-sample task discussed above, there is evidence that performance in the sequential touching task does not always converge with performance in other related tasks (Oakes, Plumert, Lansink, & Merryman, 1996). In sum, the existing empirical evidence presents some challenges for directly evaluating the prediction that across-domain differentiation should be stronger early in development relative to within-domain differentiation.

A recent study addressed these methodological challenges by using a spatial arrangement method to probe age-related changes in the domain of living kinds (Unger, Fisher, Nugent, Ventura, & MacLellan, 2016). Children were asked to arrange cards depicting animals and plants on a game board, with physical distance proportional to semantic similarity. Because the main goal of this study was to evaluate the contribution of overlapping relations to the development of semantic structure, the items used were *fully* cross-classifiable into multiple kinds of relations such that all items in the set belonged to one of three biological groups (i.e., birds, mammals, or plants) and to one of three thematic groups (i.e., aquatic, farm, or wild/zoo). The results showed that the younger children (aged 4–5 years) relied strongly on overlapping relations—judging pairs such as chicken–turkey (both birds found in farm settings) and seaweed–water lily (both plants found in water) as more strongly related than pairs that shared a single relation (e.g., chicken–eagle, chicken–lettuce)—whereas the older children showed a more graded pattern of differentiation on the basis of overlapping as well as single relations. However, the fully cross-classifiable stimulus set used in this study likely did not capture the similarity structure of most early-learned domains because it is unlikely that all items within early-learned domains are interlinked with each other by multiple overlapping relations. As such, the results reported by Unger et al. (2016) may provide a limited basis to directly evaluate the hypothesis that, early in development, within-domain differentiation should be weaker relative to across-domain distinctions—leaving this prediction still largely untested.

The current experiments aimed to fill this gap by measuring within- and across-domain differentiation across a number of early-learned domains and to do so using a single task over development that can provide graded measures of semantic structure.

### *The current experiments*

The goal of the current experiments was to *directly* examine within- and across-domain differentiation for multiple early-acquired domains over development. To directly examine age-related changes in children's semantic structure, we used a spatial arrangement method (Fisher, Godwin & Matlen, 2015; Goldstone, 1994), similar to the approach of Unger et al. (2016). We asked children to arrange items on a game board such that items that go together are placed close together; the physical distance between item pairs served as a proxy for semantic relatedness, with items judged as more similar placed closer together. This method allowed us to collect graded similarity judgments among a large number of items in a single testing session and examine both within- and across-domain differentiation for multiple domains. In Experiment 1, we concurrently assessed across- and within-domain differentiation; Experiment 2 focused on assessing only within-domain distinctions and addressed two limitations that could alternatively explain the results of Experiment 1.

To examine age-related differences in semantic structure, we compared the relative distances at which item pairs reflecting within- and across-domain distinctions were placed on the board by two age groups spanning a developmental period previously shown to undergo marked semantic

structure reorganization (e.g., Keil, 1979; Unger et al., 2016). We used age as a proxy for the amount of experience that children have accumulated, with the assumption that older children have had more experience with the domains and features relevant to the items tested.

Children's performance in the spatial arrangement task has been shown to be related to other cognitive processes thought to rely on semantic structure (Fisher, Godwin, & Matlen, 2015; Fisher, Godwin, Matlen, & Unger, 2015) and robust to the possible influence of the task-specific factors reported in prior work using other tasks such as task instructions and how the items are depicted (Fisher, Godwin, Matlen, & Unger, 2015; Unger et al., 2016)—suggesting that this task offers a valid measure of semantic structure in children.

In addition, two recent studies using the spatial arrangement method have reported pretest-to-posttest changes in the arrangements produced by the same children, with those changes being specific to children's learning experiences in a week-long summer camp at a zoo (Unger & Fisher, 2019) and a botanical garden (Vales, States, & Fisher, 2020). Specifically, children's arrangements reflected increased differentiation for the domain they experienced during the summer camp—with no such differences being observed in a control group (Unger & Fisher, 2019) or in the same children for a non-experienced domain (Vales et al., 2020).

Together, these findings suggest that developmental differences (Fisher, Godwin, & Matlen, 2015) and age-related differences (Fisher, Godwin, Matlen, & Unger, 2015; Unger et al., 2016) in semantic differentiation found using the spatial arrangement method are unlikely to be fully explained by how children of different ages may approach the task; instead, there is compelling evidence indicating that developmental and age differences observed in this task reflect differences in semantic structure.

## Experiment 1

The patterns of increased semantic differentiation over development suggested by computational modeling studies (Hills et al., 2009; Kemp & Tenenbaum, 2008; McClelland & Rogers, 2003) predict that differentiation across domains of knowledge should generally be stronger earlier in development, whereas most within-domain distinctions should emerge more slowly—and thus be less strongly represented—over the course of development. Thus, this framework predicts that both younger and older children should place pairs of items that belong to the same domain (e.g., cat and dolphin, both animals) closer together relative to pairs of items that belong to distinct domains (e.g., cat and shorts). On the other hand, this framework predicts a more graded pattern of differentiation for within-domain pairs such that older children would place items that are more strongly related within a domain (e.g., cow and chicken, often experienced together in a farm context) closer together relative to items that are less strongly related within a domain (e.g., cow and penguin), whereas younger children would show weaker or no within-domain differentiation.

To implement variation in relational strength within a domain, we created test sets in which half of the items tend to occur together in a common context (e.g., summertime). This shared theme afforded a way to differentiate items *within* a domain—that is, same-domain items that were linked by that shared context relation versus same-domain items that were not—without employing a fully cross-classifiable set of items (because not all items belonged to more than one group).

The shared theme also afforded one alternative way—relative to domains—to organize the items. Sensitivity to contextual or thematic relations has been shown early in development, and it remains an important source of relatedness throughout development (Lin & Murphy, 2001; Nelson, 1977). As such, these items linked by a common contextual relation provide an alternative way to group the items (relying on a source of relatedness to which even the younger children should be sensitive) and thus offer a strong test of the hypothesis that both age groups would show evidence of across-domain differentiation.

## Method

### Participants

A total of 50 children were recruited across two age groups; children in the younger age group were 4- and 5-year-olds ( $M = 4.5$  years,  $SD = 0.5$ ; 14 girls and 11 boys), and children in the older age group

were 7- and 8-year-olds ( $M = 7.5$  years,  $SD = 0.5$ ; 15 girls and 10 boys). The sample size per age group ( $N = 25$ ) is comparable to that in the Unger et al. (2016) study, which examined age-related differences in semantic structure using a similar task. One additional child was recruited but not included in the analyses due to experimenter error. Children were recruited from schools and preschools in Pittsburgh, PA, an urban area in the northeastern United States. The caregiver-reported racial and ethnic makeup of the sample was 40% Caucasian, 20% multiracial, and 14% African American (26% of caregivers did not report race/ethnicity). According to the 2010 U.S. census, the median income of the neighborhoods where these schools are located was \$41,899 ( $SD = 11,306$ , range = 20,431–48,363), with 26.9% of households living below the poverty line ( $SD = 17.2$ , range = 9.7–54.4). All children had normal or corrected-to-normal vision. Caregiver consent was obtained for all participants in compliance with the Carnegie Mellon University institutional review board. Children received a small gift for participating.

### Stimuli and design

There were two sets of 18 items; each item was printed on a  $5 \times 5$ -cm card with a white background. The game board on which children were asked to arrange the cards displayed a visible  $10 \times 10$  grid of 6-cm squares. The two sets are shown in Fig. 1; each set included 6 items of a domain (clothes, foods, animals, or artifacts/tools) for a total of three domains tested on each set and four domains tested across the two sets. We selected these domains because past empirical and computational work suggests that they are likely to be acquired early in life (e.g., Hills et al., 2009; Mandler & Bauer, 1988). Within each set and domain, half of the items shared an additional contextual relation (summertime or farm); these contextual relations were chosen because of their familiarity to young children. Although some domains were tested on both sets, each individual item was presented on only one set and thus was seen only once. There were two testing orders (with the order of the two sets counterbalanced) to which children were randomly assigned.

The classification of items into domains and contexts was verified in a calibration study with adults ( $N = 25$ ; see Appendix A for details). A calibration study with a separate group of children ( $N = 12$ ,  $M_{\text{age}} = 4.8$  years) using a forced-choice procedure (in which children selected a named referent out of four pictures of common objects) indicated that children could correctly identify the items by name ( $M_{\text{accuracy}} = .93$ ,  $SD = .13$ ).

### Procedure

Children sat with an experimenter in a quiet area of their school and were told that the goal of the game was to organize cards on the game board such that cards depicting items that go together are placed close together and cards depicting items that do not go together are placed far apart. While giving these task instructions, the experimenter brought her hands close together and moved them apart above the board. Prior work has shown that different instruction wordings do not result in different arrangements, with children producing comparable arrangements whether they were instructed to put together things that “are the same kinds of thing,” “go together,” or “match” (Unger et al., 2016). Children were also told that they could change the placement of the cards at any time and could take as long as they wished to arrange all cards. For each set, the experimenter laid the cards down on the table, one at a time, while labeling them (e.g., “Here is a *watermelon*”); the cards were shuffled before each participant so as to be previewed in a random order. Once all cards of a set were previewed, children were reminded of the instructions and asked to arrange the cards on the board; children were reminded of the rules again once half of the cards of a set had been arranged on the board. After all cards were placed on the board, children were asked whether they wanted to change the placement of any cards; the experimenter also clarified any cards that were not clearly placed (e.g., in between two grid cells). Once children confirmed their final arrangement, the experimenter thanked them for their help and took a photo of the board for later coding. The procedure was then repeated for the second set and was identical for both age groups. Children took no longer than 20 min to arrange the two sets.





**Fig. 1.** Items used in Experiment 1. There were two sets of 18 items. Each set included 6 items of a domain (clothes, foods, animals, or artifacts/tools), with three domains tested per set. Within each set and domain, half of the items shared an additional contextual relation: summertime (depicted at the top) or farm (depicted at the bottom).

#### Data coding

The photos of each participant's arrangements were coded by coders blind to the hypotheses of the study (see Appendix B for examples of participant-produced arrangements). A first coder used the  $10 \times 10$  grid as a coordinate plane and coded the coordinates of each card on the board; a second coder verified the coordinates of all photos (reaching 99% agreement with the first coder). We calculated pairwise distance scores for pairs of items in a set as the Euclidian distance between the two items' coordinates.

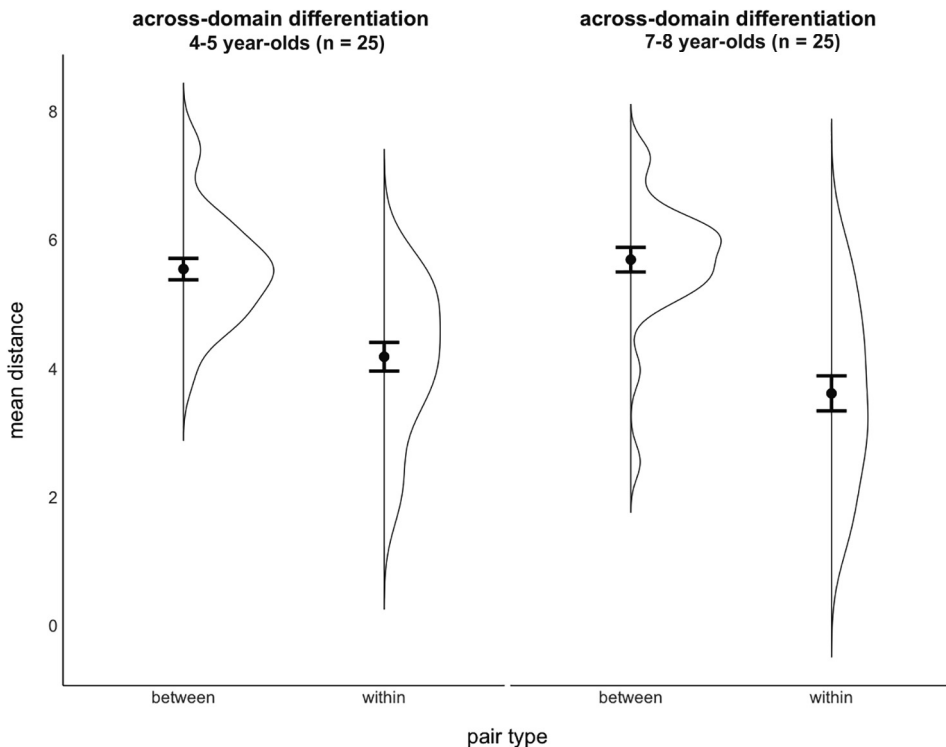
#### Results

All analyses reported in this article were conducted in the R environment (R Core Team, 2014). To examine age-related changes in children's across- and within-domain differentiation, we compared the arrangements of the two age groups by examining the distances between (a) pairs of items in either the same or different domains and (b) pairs of items in the same domain and either linked by a contextual relation or not. We used a linear mixed-effects approach to examine the effects of age group and pair type on the raw (i.e., nonaveraged) distances between each item pair using the *lmer* function from the *lme4* package (Bates, Maechler, Bolker, & Walker, 2015). Variables were centered, with categorical variables coded using effects coding. For each analysis, we fit a model with the maximal random-effects structure (Barr, Levy, Scheepers, & Tily, 2013). The *p* values reported are based on Wald tests of each model's fixed effects and were calculated using the *Anova* function from the *car* package (Fox & Weisberg, 2011). Pairwise contrasts were calculated using the *lsmeans* package (Lenth, 2016). To further explore the structure underlying children's arrangements, we also applied

average linkage hierarchical clustering (Johnson, 1967) to the pairwise distances among the items in each test set; to visualize the resulting clustering solutions, we plotted dendrograms for each age group and item set using the *hclust* function from the base *stats* package. Code, data, and procedural details are openly available through the Open Science Framework (<https://osf.io/vjkcm/>).

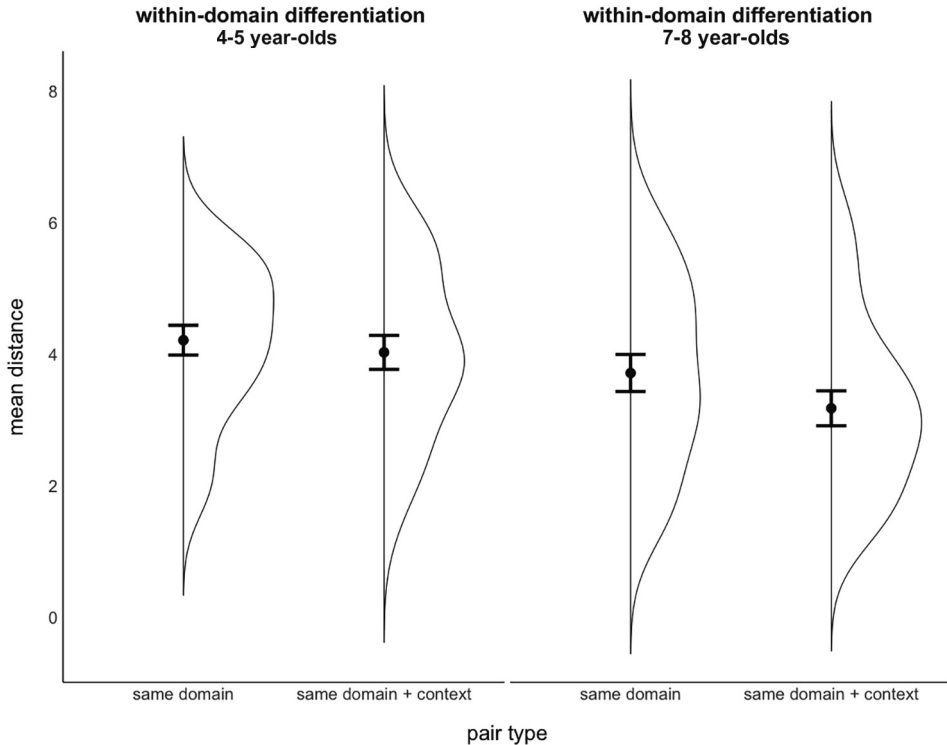
#### Distance analyses: Across-domain differentiation

Fig. 2 displays the mean distance at which pairs including 2 items from different domains (“between” pairs; e.g., cat–watermelon, an animal and a food item) or the same domain (“within” pairs; e.g., cat–dolphin, both animals) were placed on the board by the younger and older age groups. Both age groups appeared to place pairs including 2 items of the same domain closer together relative to pairs including 2 items of different domains. A model testing the effect of age (4–5 vs. 7–8 years), pair type (between vs. within), and their interaction showed that age was not a significant predictor of the distance at which items were placed on the board [ $b = -0.21$ ,  $\chi^2(1) = 0.18$ ,  $p = .67$ ]; the model included by-pair random intercepts and by-participant random intercepts and slopes for the effect of pair type. This suggests that the two age groups did not differ in how they used the space available on the board to arrange the items, a critical finding for examining age-related differences in this task. Pair type was a significant predictor of the distance at which item pairs were placed on the board [ $b = -1.72$ ,  $\chi^2(1) = 66.10$ ,  $p < .0001$ ]; pairwise contrasts confirmed that both younger children,  $t(48.44) = 4.57$ ,  $p < .0001$ , and older children,  $t(48.45) = 6.96$ ,  $p < .0001$ , significantly differentiated between the two pair types. The interaction between pair type and age was not a significant predictor of distance [ $b = -0.72$ ,  $\chi^2(1) = 2.88$ ,  $p = .09$ ]. Together, these results suggest that both younger and older children’s semantic networks are broadly differentiated into distinct domains.



**Fig. 2.** Mean distances and their distribution for pairs including 2 items from different domains (“between” pairs; e.g., cat–watermelon, an animal and a food item) or the same domain (“within” pairs; e.g., cat–dolphin). The younger age group is depicted in the left panel, and the older age group is depicted in the right panel. Error bars depict standard errors of the mean.





**Fig. 3.** Mean distances and their distributions for pairs including 2 items from the same domain that either share an additional contextual relation (“same domain + context” pairs; e.g., shorts–sunglasses) or do not (“same domain” pairs; e.g., shorts–scarf). The younger age group is depicted in the left panel, and the older age group is depicted in the right panel. Error bars depict standard errors of the mean.

#### *Distance analyses: Within-domain differentiation*

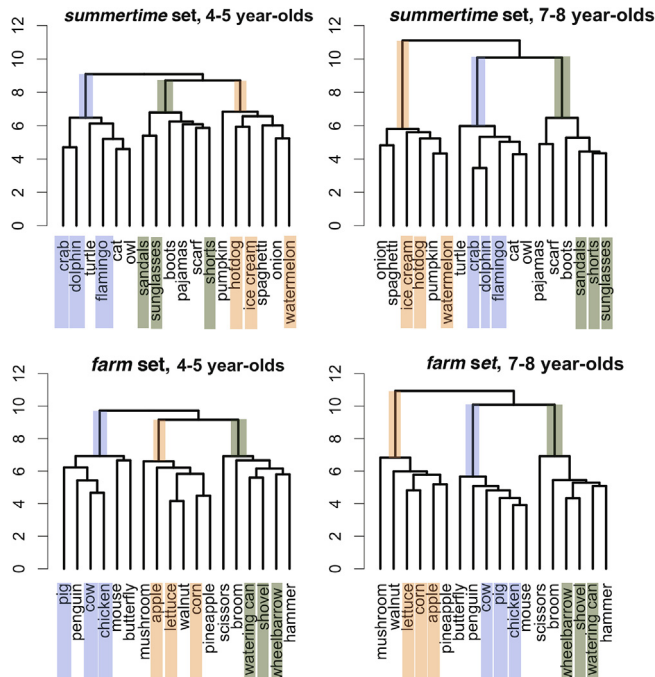
Fig. 3 displays the mean distance at which pairs including 2 items from the same domain that either shared an additional contextual relation (“same domain + context” pairs; e.g., shorts–sunglasses, clothing items likely to be worn together in the summer) or not (“same domain” pairs; e.g., shorts–scarf, clothing items that are less likely to be worn together) were placed on the board by the younger and older age groups. Older children appeared to place same-domain items that also share a contextual relation closer together relative to same-domain items that do not, whereas younger children showed a much weaker pattern of within-domain differentiation. A model testing the effect of age (4–5 vs. 7–8 years), pair type (same domain vs. same domain + context), and their interaction showed that pair type was a significant predictor of the distance at which item pairs were placed on the board [ $b = -0.36$ ,  $\chi^2(1) = 8.24$ ,  $p = .004$ ]; the model included by-pair random intercepts and by-participant random intercepts and slopes for the effect of pair type. Neither age [ $b = -0.67$ ,  $\chi^2(1) = 3.52$ ,  $p = .06$ ] nor its interaction with pair type [ $b = -0.35$ ,  $\chi^2(1) = 2.46$ ,  $p = .11$ ] were significant predictors of distance. Although the interaction between age and pair type did not reach the threshold for statistical significance, we compared the distances at which the two pair types were placed by each age group separately because we had an a priori hypothesis for age-related differences (see Hancock & Klockars, 1996, Howell, 2010, and Ryan, 1959, for discussions of why it is appropriate to conduct these analyses in the absence of an interaction because requiring both the overall model and follow-up contrasts to reach significance can increase the odds of a false negative); these pairwise contrasts showed that whereas older children significantly differentiated within domains,  $t(55.41) = 3.19$ ,  $p = .002$ , younger children showed no evidence of within-domain differentiation,  $t(55.38) = 1.10$ ,  $p = .28$ .

### Hierarchical clustering analyses

To capture the latent structure underlying children's arrangements, we applied average linkage hierarchical clustering (Johnson, 1967) to the average pairwise distances among the items in each test set and plotted dendrograms for each age group and item set to visualize the resulting clustering solutions. This agglomerative method starts by assigning each item to a single node and progressively merges the nodes into clusters based on their similarities. A dendrogram depicts this sequence of merging, providing a description of the structure of the data based on the relative similarities among the items. The height at which any two objects are linked together in a dendrogram is proportional to their similarity, with items within the same cluster being more similar to each other relative to items assigned to other clusters. By providing a description of the similarity structure of a dataset, this analytical approach offers an additional unsupervised way to examine the structure of children's arrangements in this task.

Fig. 4 depicts the dendrograms for each age group and item set tested in Experiment 1. As can be seen, for each set and age group, the three clusters at the highest level correspond to the three domains tested in each testing set. Bootstrapped  $p$  values calculated using the *pvclust* function (Suzuki & Shimodaira, 2006) confirmed that the three highest-level clusters were strongly supported by the data for both sets and age groups (all  $ps < .05$ ). This suggests that both age groups differentiated among the domains tested in this experiment.

Within each of these domains, items that share an additional contextual relation seem more likely to be linked at a lower level in the older age group than in the younger age group. A two-sample  $t$  test comparing the heights at which within-domain items that share a contextual relation become merged under the same cluster confirmed that the average height was significantly lower for the older age



**Fig. 4.** Dendrograms depicting the clustering solutions for each age group (left vs. right panels) and item set (top vs. bottom panels) in Experiment 1. The level at which all items from each domain tested are joined under the same cluster is indicated by a colored arc; nodes representing items within a domain that share a context relation are highlighted in the same color to differentiate them from within-domain items that do not (not highlighted). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

group ( $M = 5.21$ ,  $SD = 0.50$ ) relative to the younger age group ( $M = 6.41$ ,  $SD = 0.25$ ),  $t(10) = 5.26$ ,  $p = .0004$ , but no such difference was seen for within-domain items that do not share an additional relation,  $t(10) = 1.65$ ,  $p = .13$ —suggesting a larger degree of within-domain differentiation for the older age group. In sum, the structure of children's arrangements captured by the hierarchical clustering solutions converges with the results from the distance analyses.

## Discussion

Taken together, the results of Experiment 1 are consistent with the hypothesis that while both younger and older children's semantic networks are differentiated into broad domains of knowledge, within-domain differentiation is slower to emerge—with the younger children showing no evidence of within-domain differentiation in this experiment. Importantly, there were no effects of age on the overall distance at which items were placed in the board, suggesting that the two age groups did not use the available space in different ways. These results are consistent with the predictions by the framework discussed in the Introduction, suggesting that over development, and likely as a result of experience, children are accruing knowledge about items from distinct domains—with the similarity structure that characterizes most across-domain distinctions generally supporting its earlier differentiation relative to within-domain distinctions.

However, the results from Experiment 1 present two main limitations. First, the interaction between pair type and age group was not statistically significant for the within-domain distance analyses; although the planned contrasts suggest that only the older age group showed evidence of within-domain differentiation, the lack of a significant interaction raises the possibility that the current experiment was underpowered to examine within-domain differentiation. If this were so, then the apparent lack of within-domain differentiation for the younger children might not reflect slower emergence of within-domain differentiation but instead might be the result of a false negative due to low statistical power. Given that the sample size per age group in this experiment was comparable to—and in fact larger than—that in a prior study examining age-related differences across pair types using the same experimental paradigm and testing a similar age range (Unger et al., 2016), it seems unlikely that the number of participants in Experiment 1 would be the main factor limiting statistical power. However, it is possible that the *number of pairs* used to examine within-domain distinctions may have been the limiting factor. Specifically, whereas there were 36 “same domain” pairs per set, there were only 9 “same domain + context” pairs per set. Because both the number of participants and the number of measurements per participant contribute to statistical power (Carvalho, Braithwaite, de Leeuw, Motz, & Goldstone, 2016; DeBolt, Rhemtulla, & Oakes, 2020; Forrester, 2015), and the sample size was unlikely to be the main factor limiting statistical power, it is possible that Experiment 1 included too few “same domain + context” pairs to examine age-related differences in within-domain differentiation.

In addition, it is also possible that the multiple domains and relations presented in each set—and thus the need to attend to multiple ways in which the items could be arranged—rendered within-domain distinctions less noticeable for the younger age group. If this were so, then the results of Experiment 1 might not reflect an absolute lack of within-domain differentiation in the younger age group but instead might be driven, at least in part, by task demands. This possibility is consistent with recent empirical evidence showing that children as young as 2 years are sensitive to multiple relations among items (Arias-Trejo & Plunkett, 2009; Unger et al., 2016), suggesting that young children could have differentiated between same-domain items that shared an additional context relation and same-domain items that did not. It is important to note that the younger children showing evidence of across-domain (but not within-domain) differentiation when presented with items that could be differentiated at multiple levels is fully consistent with across-domain distinctions being more strongly represented early in development. Nonetheless, it is important to obtain converging evidence to the observed age differences in within-domain differentiation and ensure that they are not merely due to task demands.

The goals of Experiment 2 were twofold. First, Experiment 2 aimed to address the two limitations of Experiment 1 discussed above. Specifically, in Experiment 1 we included a larger number of same-domain items that were linked by an additional relation and reduced the number of relations pre-

sented on each testing set by testing only one domain at a time. Second, Experiment 2 was designed to further test the hypothesis that there should be age-related differences in within-domain differentiation, as predicted by the framework discussed in the Introduction. Toward these two goals, we asked children to arrange sets of same-domain items with only one primary differentiating relation, which included a comparable number of each type of item pair.

## Experiment 2

The main goal of Experiment 2 was to provide an additional test of the hypothesis that there should be age-related differences in within-domain differentiation by increasing the number of measurements for each type of within-domain pair and reducing potential task demands that can alternatively explain the results of Experiment 1. Thus, in Experiment 2 we examined children's within-domain differentiation by testing only one domain at a time and including a larger number of same-domain items that were linked by an additional relation. Specifically, within each set we presented children with items of a single domain (e.g., only food items) that could be differentiated based on a single relation (e.g., breakfast vs. dinner foods); to prevent additional relations from driving children's performance, we also ensured that all items within that domain belonged to a similar category (e.g., that all food items in a test set were grains). This approach minimized additional apparent ways in which items could be organized—a potential task demand that could have contributed to children's performance in Experiment 1—and resulted in a larger number of same-domain items sharing an additional relation relative to Experiment 1.

Thus, if the results of Experiment 1 are due to a small number of pairs including same-domain items that share an additional relation, as well as to the concurrent testing of across- and within-domain relations, we expected that in Experiment 2 even the younger children would show evidence of within-domain differentiation. In addition, by the predictions of the framework discussed in the Introduction (and consistent with the findings of Experiment 1), we would still expect older children to show *larger* within-domain differentiation relative to younger children.

## Method

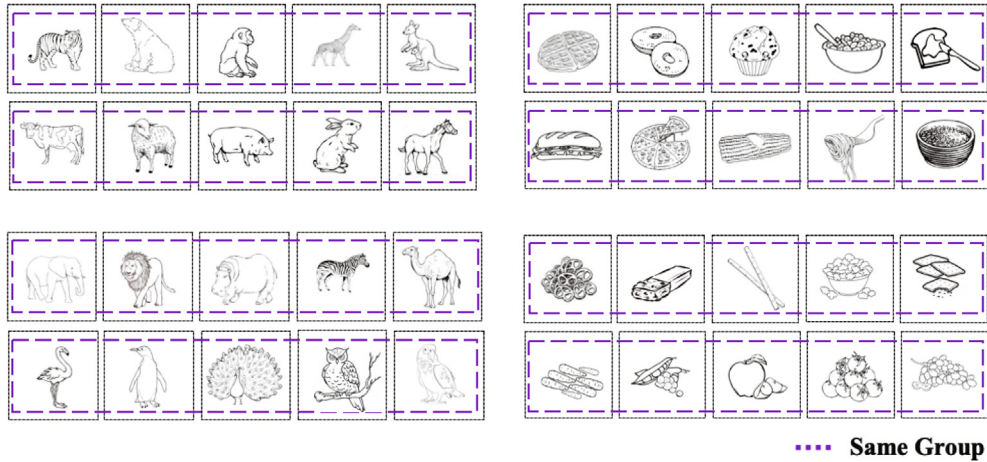
### Participants

Recruitment and testing procedures were identical to those in Experiment 1. A total of 50 children were recruited across two age groups: 4- and 5-year-olds ( $M = 4.8$  years,  $SD = 0.4$ ; 15 girls and 10 boys) and 7- and 8-year-olds ( $M = 7.5$  years,  $SD = 0.5$ ; 10 girls and 15 boys); none of these children participated in Experiment 1. Two additional children were recruited but not included in the analyses due to experimenter error. The caregiver-reported racial and ethnic makeup of the sample was 72% Caucasian, 10% African American, 2% multiracial, and 2% Asian (14% of caregivers did not report race/ethnicity). According to the 2010 U.S. census, the median income of the neighborhoods where these schools are located was \$43,366 ( $SD = 17,420$ , range = 30,273–75,301), with 17.6% of households living below the poverty line ( $SD = 9.9$ , range = 2.9–31.2).

### Stimuli, design, procedure, and data coding

There were four testing sets of 10 items each. Items within the same set belonged to a single domain (e.g., food) and to one of two within-domain groups (e.g., breakfast vs. dinner foods), with 5 items per group. Items within each testing set were of similar kind (e.g., all foods in a testing set were grains). Fulfilling the goal of having a larger number of same-domain items sharing an additional relation relative to Experiment 1, each testing set included 25 “same domain” pairs and 20 “same domain + group” pairs.

To extend the findings of Experiment 1, across sets we tested two kinds of within-domain distinctions (contextual and taxonomic groupings); in Appendix C, we report exploratory analysis for each grouping type separately. The four sets were wild versus farm animals (all mammals), breakfast versus dinner foods (all grains), birds versus mammals (all wild animals), and grains versus fruits (all snack foods). The items used are displayed in Fig. 5. The classification of items into within-domain groups



**Fig. 5.** Items used in Experiment 2. There were four sets of 10 items each. Items within a set belonged to a single domain (i.e., animals [left] or foods [right]) and group (i.e., mammals, grains, wild animals, or snack foods). To implement within-domain distinctions, half of the sets implemented a contextual distinction (depicted at the top: wild vs. farm animals; breakfast vs. dinner foods) and the other half implemented a taxonomic distinction (depicted at the bottom: birds vs. mammals; grains vs. fruits/veggies).

and categories was verified in a study with adults ( $N = 25$ ; see Appendix A). A separate group of children ( $N = 13$ ,  $M_{\text{age}} = 4.9$  years) correctly identified the items by name ( $M_{\text{accuracy}} = .98$ ,  $SD = .02$ ). Each child completed a total of 4 trials, with 1 trial for each testing set; examples of participant-produced arrangements are available in Appendix B. All other design, procedural, and coding details were identical to those in Experiment 1.

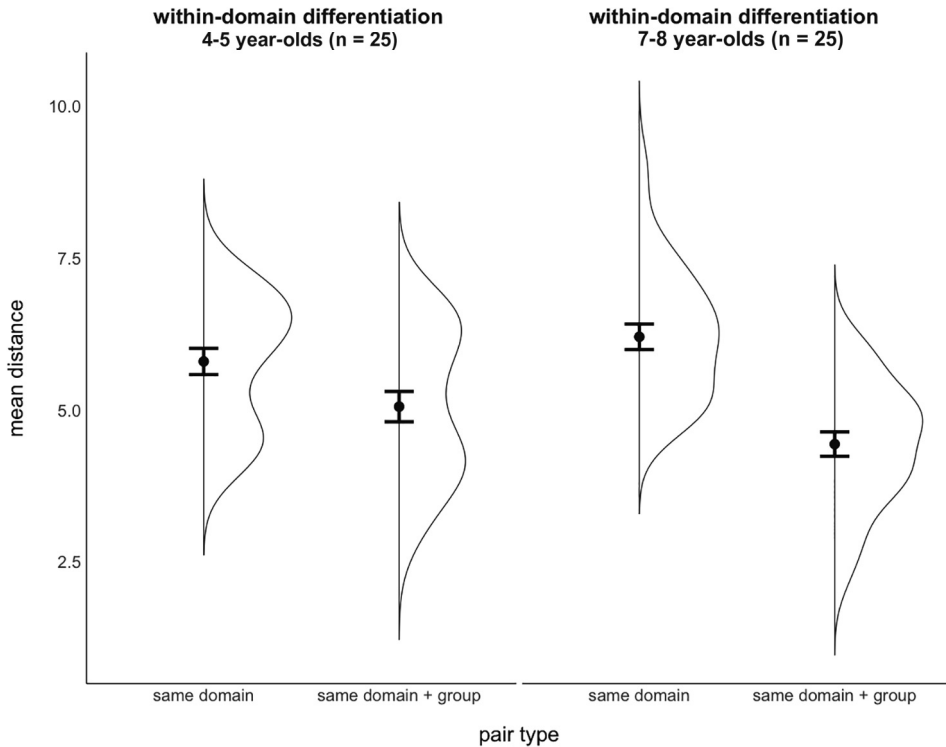
## Results

### Distance analyses

Fig. 6 displays the mean distance between pairs including 2 items from the same domain that either share an additional relation (“same domain + group” pairs; e.g., elephant–lion, both animals and mammals) or do not (“same domain” pairs; e.g., elephant–flamingo, both animals, one being a mammal and the other being a bird). Although both age groups appeared to place same-domain items that share an additional relation closer together relative to same-domain items that do not, there also seemed to be stronger differentiation in the older age group. A model testing the effect of age (4–5 vs. 7–8 years), pair type (same domain vs. same domain + group), and their interaction showed that whereas pair type [ $b = -1.26$ ,  $\chi^2(1) = 45.38$ ,  $p < .0001$ ] was a significant predictor of the distance at which item pairs were placed on the board, age was not [ $b = -0.11$ ,  $\chi^2(1) = 0.05$ ,  $p = .82$ ]; the model included by-pair random intercepts and by-participant random intercepts and slopes for the effect of pair type. The interaction between pair type and age was a significant predictor of distance [ $b = -1.02$ ,  $\chi^2(1) = 8.49$ ,  $p = .004$ ], suggesting that older children differentiated items within a domain to a larger extent than younger children. Pairwise contrasts showed that both younger children,  $t(53.78) = 2.87$ ,  $p = .006$ , and older children,  $t(54.96) = 6.94$ ,  $p < .0001$ , differentiated between the two pair types.

### Hierarchical clustering analyses

Fig. 7 depicts the dendrograms for each age group and set tested in Experiment 2. As can be seen, within each domain tested, items that belong to the same within-domain group appeared more likely to be linked under the same cluster in the older age group than in the younger age group. To explore these patterns, we compared the average heights at which the items that belong to the same group become merged under the same cluster across the two age groups for each set; for those sets in which the 5 items of the same group did not cluster together, we used the height at which most of the items



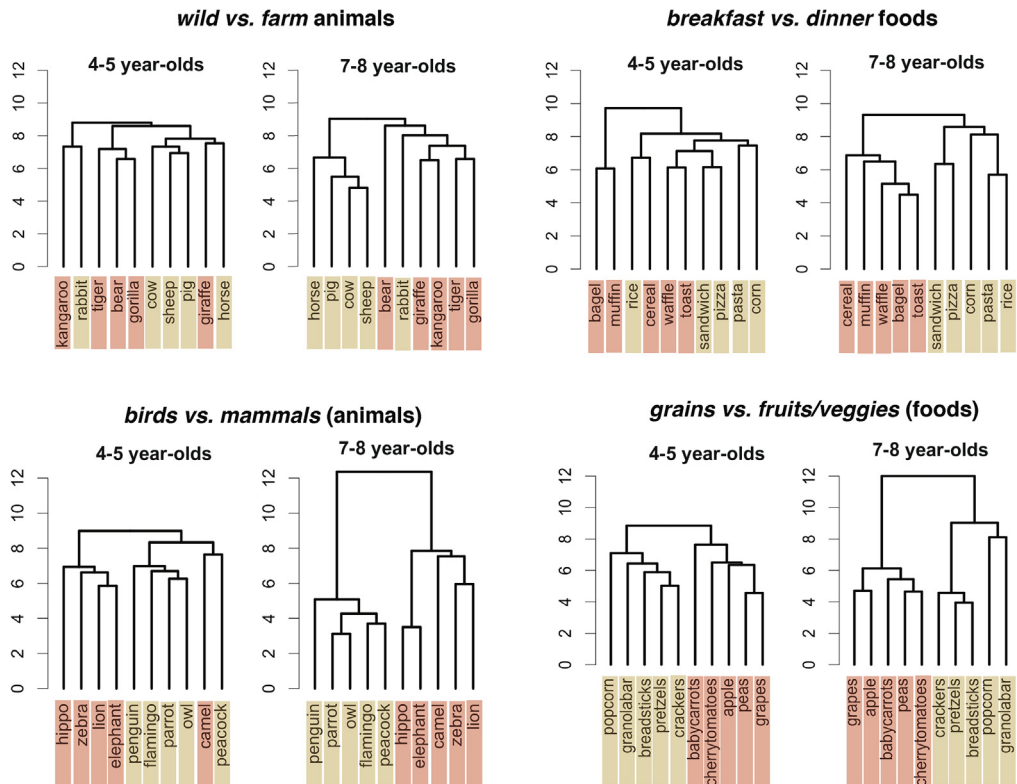
**Fig. 6.** Mean distances and their distributions for pairs including 2 items from the same domain that either share an additional relation (“same domain + group” pairs; e.g., elephant–lion) or do not (“same domain” pairs; e.g., elephant–flamingo). The younger age group is depicted in the left panel, and the older age group is depicted in the right panel. Error bars depict standard errors of the mean.

that belong to the same group clustered together. A two-sample  $t$  test confirmed that the average height at which items from the same group become merged under the same cluster was significantly lower for the older age group ( $M = 7.34$ ,  $SD = 1.50$ ) relative to the younger age group ( $M = 8.72$ ,  $SD = 0.92$ ),  $t(16) = 2.22$ ,  $p = .04$ —suggesting a larger degree of within-domain differentiation for the older age group. In sum, the structure of children’s arrangements captured by the hierarchical clustering solutions converges with the results from the distance analyses.

## Discussion

The results of Experiment 2 converge with those of Experiment 1 by showing age-related increases in within-domain differentiation such that older children differentiated items within a domain to a larger extent than younger children. However, whereas in Experiment 1 the younger age group showed no evidence of within-domain differentiation, in Experiment 2 even the younger children were able to differentiate within the domains presented—albeit less strongly than the older children. Thus, these results suggest that the lack of within-domain differentiation found in the younger age group in Experiment 1 could be due to (a) the use of a smaller number of same-domain pairs that shared an additional relation, resulting in lower statistical power to detect within-domain differentiation, as well as to (b) the concurrent testing of multiple domains and within-domain relations, which may have rendered within-domain distinctions less evident for younger children.





**Fig. 7.** Dendrograms depicting the clustering solutions for each item set and age group (left vs. right within a panel). Nodes representing items within a domain that share an additional relation are highlighted in the same color. (For interpretation of the reference to color in this figure legend, the reader is referred to the Web version of this article.)

## General discussion

Organized semantic representations encoding across- and within-domain distinctions are a hallmark of mature cognition, and understanding how they change with experience and learning is a key endeavor of cognitive theories. The results reported here provide the first direct evidence that, early in development, across-domain distinctions are generally more strongly represented relative to within-domain distinctions. In Experiment 1, only older children reliably differentiated items within a domain, but both age groups differentiated among items belonging to distinct domains. The findings from Experiment 2 showed that, when presented only with same-domain items, even the younger age group showed some evidence of within-domain differentiation—but still to a lesser degree than older children. In other words, whereas across-domain distinctions seem to be easily accessible to—and thus strongly represented by—both age groups, within-domain distinctions (although not fully inaccessible) seem to be less strongly represented by the younger age group. Together, these findings show that, over development, the early aggregation of items into domains is generally followed by a more protracted splitting of those items according to within-domain distinctions.

This pattern of results provides direct evidence in support of a key prediction by the theoretical framework presented in the Introduction. Specifically, because the properties that discriminate items within a domain tend to covary with the properties that are shared across all items of that domain, this framework predicts that early in development within-domain differentiation should generally

be weaker relative to across-domain differentiation. By providing direct evidence to this prediction, the current results support the idea that structured semantic representations emerge from gradually learning about the properties of entities in the world and from capitalizing on the patterns of covariation among those properties.

The current results are in line with a number of recent findings documenting the role of repeated experiences with items of a domain in increasing both across-domain differentiation (Vales et al., 2020) and within-domain differentiation (Badger & Shapiro, 2019; Unger & Fisher, 2019; Vales et al., 2020) and add to this literature by examining this hypothesis in a cross-sectional sample. These results are also broadly consistent with recent work suggesting that semantic networks become increasingly connected over the life span (Dubossarsky, De Deyne, & Hills, 2017), suggesting that experience-based learning likely continues to shape semantic structure over the life span.

The results reported here could be taken to conflict with the results reported by Unger et al. (2016), a recent study addressing many of the methodological challenges from prior studies examining semantic structure in childhood. Specifically, the results of Unger et al. suggested that within-domain distinctions (e.g., distinguishing farm animals from wild animals, distinguishing farm plants from water plants) were more strongly represented earlier in development relative to across-domain distinctions (e.g., animals vs. plants) —the opposite pattern of results observed in the current experiments. However, these apparent differences can be reconciled by considering the similarity structure of the specific items tested by Unger et al. (2016). Specifically, the item set used by Unger et al. was fully cross-classifiable and presented same-domain items that could be distinguished based on nonoverlapping relations as well as different domain items that shared overlapping relations. Although suitable to examine the main question of the Unger et al. study, according to the framework discussed in the Introduction, such similarity structure would be expected to facilitate within-domain differentiation and to hinder across-domain differentiation *given the specific items tested*. This is because the strength of differentiation is expected to depend on the similarity patterns of the specific items tested—same-domain groupings that rely on multiple shared relations should be more easily acquired, and across-domain items that share overlapping relations should be more slowly differentiated. As such, both the results of Unger et al. (2016) and the current results can be accounted for by a framework that considers the similarity patterns of the specific items tested. Future work can more systematically test the predictions of this framework by examining whether the strength of semantic differentiation can be predicted by the specific patterns of feature covariation among items of early-acquired domains; for example, distinctions between items with fewer consistently covarying properties should be acquired earlier in development. Future experimental work can also manipulate the similarity structure of experienced items to provide converging evidence to this framework.

The framework suggested by the results of the computational modeling studies discussed in the Introduction proposes that experiencing the properties of items in the world is central to the emergence of structured semantic representations (Hills et al., 2009; Kemp & Tenenbaum, 2008; McClelland & Rogers, 2003). Although they provide a mechanistic framework for conceptualizing how semantic structure changes with experience and learning, these models are mostly agnostic about the specific cognitive mechanisms by which these properties are attended to in the moment, processed over multiple time scales, and ultimately learned. Furthermore, all item properties are often treated as contributing equally to semantic differentiation, but it is likely that different properties contribute to semantic differentiation to different extents (see Hills et al., 2009; Peters & Borovsky, 2019). A deeper understanding of how item properties are learned and how different kinds of properties contribute to semantic differentiation is key to establishing more comprehensive theories of semantic development.

The literature on semantic structure has also emphasized the key role that semantic knowledge plays in supporting other cognitive processes, including reasoning, word learning, and language processing (e.g., Borovsky et al., 2016; Colunga & Sims, 2017; Federmeier & Kutas, 1999; Fisher, Godwin, & Matlen, 2015; Gobbo & Chi, 1986). However, the specific mechanisms by which organized semantic representations support cognition remain less well understood, particularly early

in development. A number of recent studies suggest that the momentary coactivation of semantically related items may facilitate the integration and use of incoming information that overlaps with the existing semantic structure (Borovsky & Peters, 2019; Hutchinson & Turk-Browne, 2012; Vales & Fisher, 2019), making the coactivation of related items a plausible candidate for how semantic structure supports other cognitive processes. Future developmental experiments should directly test these predictions so as to better understand how children's semantic representations support their learning of new information.

Finally, the current results also make a methodological contribution. Across the two experiments reported here, whether younger children showed evidence of within-domain differentiation depended on whether multiple levels of semantic structure were assessed simultaneously as well as on the number of items belonging to each within-domain grouping; specifically, younger children showed evidence of differentiating within domains of knowledge only when tested on one domain at a time and with a large number of same-domain pairs that also shared an additional relation. Future work examining semantic differentiation in young children may want to consider whether to concurrently or independently test multiple levels of differentiation and how many items to include. On this point, it should also be noted that the specific within-domain items tested in Experiments 1 and 2 were not identical, which could have contributed to whether children could discriminate items within a domain. The difference in items seems unlikely to fully explain the differences between the two experiments on three grounds. First, both experiments tested within-domain distinctions with which children should be familiar (Arias-Trejo & Plunkett, 2009; Nelson, 1977; Unger et al., 2016), but only in Experiment 2 did the younger children show evidence of within-domain differentiation. Second, all the domains tested in both experiments are likely to be acquired early in development (Hills et al., 2009; Mandler & Bauer, 1988) and thus should be familiar to young children. Third, the items in both experiments were shown to be recognizable by—and thus familiar to—young children in calibration studies. Together, these reasons make it unlikely that the differences in the specific items used could account for the observed difference in within-domain differentiation between the two experiments and underscore the importance of considering whether to concurrently or independently test multiple levels of differentiation and how many items to include when assessing semantic structure in young children.

## Conclusions

The results of the experiments reported here show that, early in development, children's semantic networks more strongly represent distinctions across broad domains of knowledge, with within-domain differentiation emerging more slowly over the course of development. These results are the first to directly test this prediction derived from existing computational modeling studies and are consistent with a mechanistic framework that emphasizes the exploitation of patterns of covariation in the acquisition of structured semantic representations.

## Acknowledgments

We would like to thank Jiwon Ban, Rea Isaac, Cecilia Mastrogiacomio, Oceann Stanely, and Anna Vande Velde for their help with stimuli creation, data collection, and coding; our community partners and the children who participated in these studies for making this research possible; and Nynke Niezink for statistical advice. This research was partially supported by the James S. McDonnell Foundation (award 220020401 to A.V.F.), the National Science Foundation (award 1918259 to A.V.F. and C.V.), and a Carnegie Mellon University Simon Initiative Seed Grant to A.V.F. and C.V. The Institute of Education Sciences, U.S. Department of Education (award R305B150008) supported P.S. The opinions expressed are those of the authors and do not represent the views of the funding sources.

## Appendix A

### Stimuli calibration

A total of 25 adults participated in a calibration study to verify the assignment of items into domains and contexts (for the stimulus set used in Experiment 1) and into groups and categories (for the stimulus set used in Experiment 2). Participants were undergraduate students at Carnegie Mellon University and participated in exchange for course credit. All participants were native speakers of English. Consent was obtained for all participants.

#### *Experiment 1 stimulus set*

To verify the assignment of items into contexts, participants were asked to rate, on an 11-point scale, how strongly related with either the *summer* or *farm* context each item on that set was. For the *summer* item set, participants rated the items intended to be strongly associated with that context as more strongly associated with the summer context ( $M = 8.8$ ,  $SD = 1.3$ ) relative to the remaining items of that set ( $M = 3.7$ ,  $SD = 1.0$ ),  $t(16) = 9.32$ ,  $p < .0001$ . For the *farm* item set, participants also rated the items intended to be strongly associated with that context as more strongly associated with the farm context ( $M = 9.5$ ,  $SD = 0.5$ ) relative to the remaining items of that set ( $M = 5.3$ ,  $SD = 1.7$ ),  $t(16) = 7.11$ ,  $p < .0001$ .

To verify the assignment of items into domains, participants were asked to select, for each item, the domain that best described that item from the following four options: animals, artifacts/tools, foods, and clothes. All items were more likely to be assigned to the intended domain than to any of the other domains, and the likelihood of domain assignment was higher than what would be expected by chance (.25) for all four domains (all  $ts > 15$ , all  $ps < .0001$ ).

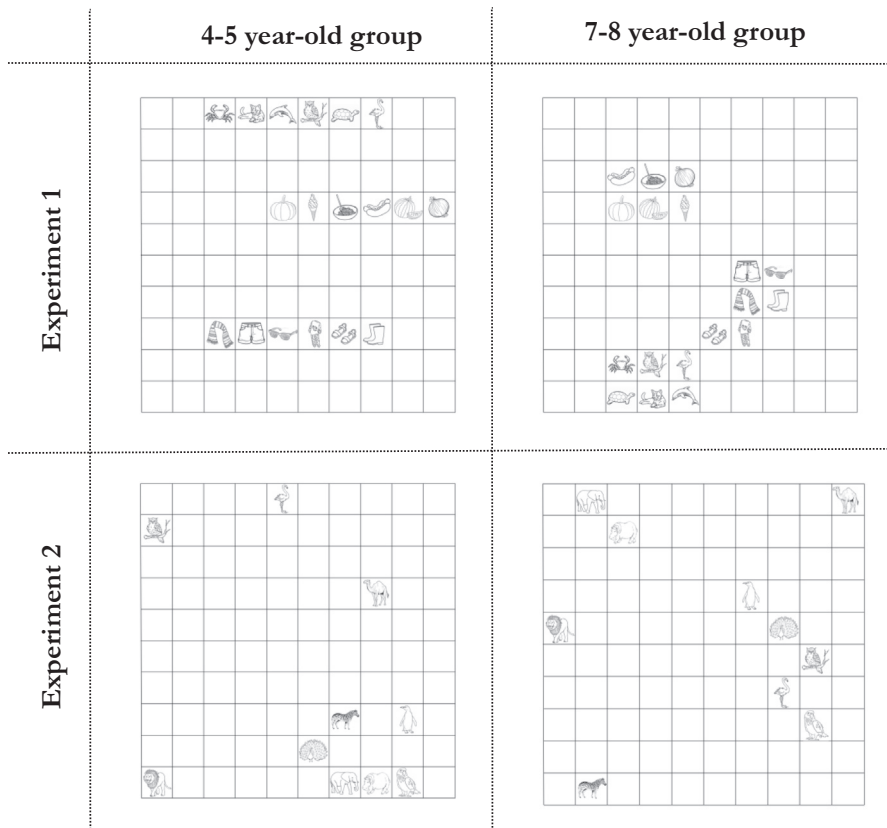
#### *Experiment 2 stimulus set*

To verify the assignment of items into within-domain groups, participants were asked to select, for each item of a set, the group that best described that item. Each set targeted one of the following within-domain distinctions: farm versus wild animals; breakfast versus dinner foods; birds versus mammals; veggies/fruits versus bread/grains. All items but one (rabbit) were more likely to be assigned to the intended group than to any of the other groups; the likelihood of group assignment was higher than what would be expected by chance (.50) for all groups (all  $ts > 2.37$ , all  $ps < .05$ ).

To verify that the items were perceived as belonging to the intended category (i.e., mammals, wild animals, grains, and snack foods), participants were asked to select, for each item of a set, the category that best described that item. For participants to select between two alternatives, each of the intended categories was contrasted with another category in the same domain. Specifically, for each set, participants were asked to select between one of the following categories: mammals versus birds; wild versus domestic animals; grains versus greens; snacks versus beverages. All items but one ("parrot") were more likely to be assigned to the intended category than to the contrasting category; the likelihood of category assignment was higher than what would be expected by chance (.50) for all target categories (all  $ts > 6.49$ , all  $ps < .001$ ).

## Appendix B

See Fig. B1 for Examples of participant-produced arrangements



**Fig. B1.** Example of a participant-produced arrangement for each age group (left vs. right panels) and experiment (top vs. bottom panels).

## Appendix C

### Exploratory analyses (Experiment 2): Within-domain differentiation per grouping type

In Experiment 2 reported in the main article, children were asked to arrange four same-domain testing sets. Two tests in each domain targeted contextual distinctions (wild vs. farm animals; breakfast vs. dinner foods), and the other two targeted taxonomic distinctions (birds vs. mammals; grains vs. fruits/veggies). Here we report additional analyses examining the degree to which within-domain

**Table C1**

Within-domain differentiation: Coefficient estimates, Wald chi-square tests, and significance level for all predictors.

| Predictor                                       | Coefficient | $\chi^2$ | <i>p</i> Value   |
|---|-------------|----------|------------------|
| Age group (4–5 vs. 7–8 years)                   | –0.11       | 0.07     | .80              |
| Pair type (same domain vs. same domain + group) | –1.26       | 47.57    | <b>&lt;.0001</b> |
| Grouping type (contextual vs. taxonomic)        | –0.19       | 0.79     | .37              |
| Age Group × Pair Type                           | –1.02       | 8.49     | <b>.004</b>      |
| Age Group × Grouping Type                       | 0.40        | 2.62     | .10              |
| Pair Type × Grouping Type                       | –1.09       | 50.84    | <b>&lt;.0001</b> |
| Age Group × Pair Type × Grouping Type           | –0.75       | 11.69    | <b>.0006</b>     |

Note. Significant *p* values are bolded.

distinctions varied as a function of the grouping presented (contextual vs. taxonomic). We note that we did not select the items in the sets to control along other dimensions that likely also matter for differentiation; for example, there likely were systematic perceptual differences between items that were greater for some sets than for others (e.g., mammals and birds may be more perceptually distinct relative to farm vs. wild mammals). As such, we regard these analyses as exploratory.

We built a model examining the contributions of pair type, age group, grouping type (contextual vs. taxonomic), and their interactions in the same way as described for Experiment 2 in the main article; the results of this model are displayed in Table C1. The main takeaway is that the three-way interaction among pair type, age group, and grouping type was a significant predictor of the distance at which within-domain item pairs were placed on the board. Pairwise contrasts showed that the older age group significantly differentiated both contextual sets,  $t(66.35) = 3.85$ ,  $p = .0003$ , and taxonomic sets,  $t(66.35) = 9.31$ ,  $p < .0001$ , whereas the younger age group significantly differentiated taxonomic sets,  $t(65.98) = 4.02$ ,  $p = .0002$ , but not contextual sets,  $t(65.98) = 1.42$ ,  $p = .16$ .

The post hoc nature of these findings does not warrant strong conclusions; nevertheless, it is possible that specific domains of knowledge, and specific items within those domains of knowledge, become differentiated at different rates as a result of experience and learning. The mechanistic framework discussed in the Introduction of the main article offers one way to measure the degree to which specific items should differentiate over development by considering the degree of consistent covariation among item features; future work can more directly test these possibilities using stimulus sets appropriate for examining these hypotheses.

## References

- Arias-Trejo, N., & Plunkett, K. (2009). Lexical-semantic priming effects during infancy. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 364, 3633–3647.
- Badger, J. R., & Shapiro, L. R. (2019). We're going to the zoo: Interactive educational activities with animals boost category-based induction in children. *Cognitive Development*, 49, 1–10.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1–48.
- Bjorklund, D. F., & Jacobs, J. W. (1985). Associative and categorical processes in children's memory: The role of automaticity in the development of organization in free recall. *Journal of Experimental Child Psychology*, 39, 599–617.
- Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Semantic structure in vocabulary knowledge interacts with lexical and sentence processing in infancy. *Child Development*, 87, 1893–1908.
- Borovsky, A., & Peters, R. E. (2019). Vocabulary size and structure affects real-time lexical recognition in 18-month-olds. *PLoS ONE*, 14(7), e219290.
- Boster, J. S., & Johnson, J. C. (1989). Form or function: A comparison of expert and novice judgments of similarity among fish. *American Anthropologist*, 91, 866–889.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT Press.
- Carvalho, P. F., Braithwaite, D. W., de Leeuw, J. R., Motz, B. A., & Goldstone, R. L. (2016). An in vivo study of self-regulated study sequencing in introductory psychology courses. *PLoS ONE*, 11(3), e152115.
- Clark, E. V. (1973). What's in a word? On the child's acquisition of semantics in his first language. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 65–110). New York: Academic Press.
- Coley, J. D. (2012). Where the wild things are: Informal experience and ecological reasoning. *Child Development*, 83, 992–1006.
- Colunga, E., & Sims, C. E. (2017). Not only size matters: Early-talker and late-talker vocabularies support different word-learning biases in babies and networks. *Cognitive Science*, 41(Suppl. 1), 73–95.
- Cree, G. S., & McRae, K. (2003). Analyzing the factors underlying the structure and computation of the meaning of chipmunk, cherry, chisel, cheese, and cello (and many other such concrete nouns). *Journal of Experimental Psychology: General*, 132, 163–201.
- Crowe, S. J., & Prescott, T. J. (2003). Continuity and change in the development of category structure: Insights from the semantic fluency task. *International Journal of Behavioral Development*, 27, 467–479.
- DeBolt, M. C., Rhemtulla, M., & Oakes, L. M. (2020). Robust data and power in infant research: A case study of the effect of number of infants and number of trials in visual preference procedures. *Infancy*, 25(4), 393–419.
- Dubossarsky, H., De Deyne, S., & Hills, T. T. (2017). Quantifying the structure of free association networks across the life span. *Developmental Psychology*, 53, 1560–1570.
- Federmeier, K. D., & Kutas, M. (1999). A rose by any other name: Long-term memory structure and sentence processing. *Journal of Memory and Language*, 41, 469–495.
- Fisher, A. V., Godwin, E. K., & Matlen, B. (2015). Development of inductive generalization with familiar categories. *Psychonomic Bulletin & Review*, 22, 1149–1173.
- Fisher, A. V., Godwin, K. E., Matlen, B. J., & Unger, L. (2015). Development of category-based induction and semantic knowledge. *Child Development*, 86, 48–62.



- Forrester, S. E. (2015). Selecting the number of trials in experimental biomechanics studies. *International Biomechanics*, 2(1), 62–72.
- Fox, J., & Weisberg, S. (2011). *An R companion to applied regression*. Thousand Oaks, CA: Sage.
- Gobbo, C., & Chi, M. (1986). How knowledge is structured and used by expert and novice children. *Cognitive Development*, 1, 221–237.
- Goldstone, R. (1994). An efficient method for obtaining similarity data. *Behavior Research Methods, Instruments, & Computers*, 26, 381–386.
- Hancock, G. R., & Klockars, A. J. (1996). The quest for alpha: Developments in multiple comparison procedures in the quarter century since Games (1971). *Review of Educational Research*, 66, 269–306.
- Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, 112, 381–396.
- Howell, D. C. (2010). *Statistical methods for psychology* (8th ed.). Belmont, CA: Cengage Wadsworth.
- Hutchinson, J. B., & Turk-Browne, N. B. (2012). Memory-guided attention: Control from multiple memory systems. *Trends in Cognitive Sciences*, 16, 576–579.
- Inagaki, K. (1990). The effects of raising animals on children's biological knowledge. *British Journal of Developmental Psychology*, 8, 119–129.
- Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32, 241–254.
- Kaduk, K., Bakker, M., Juvrud, J., Gredebäck, G., Westermann, G., Lunn, J., & Reid, V. M. (2016). Semantic processing of actions at 9 months is linked to language proficiency at 9 and 18 months. *Journal of Experimental Child Psychology*, 151, 96–108.
- Keil, F. C. (1979). *Semantic and conceptual development: An ontological perspective*. Cambridge, MA: Harvard University Press.
- Kemp, C., & Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 10687–10692.
- Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software*, 69(1), 1–33.
- Lin, E. L., & Murphy, G. L. (2001). Thematic relations in adults' concepts. *Journal of Experimental Psychology: General*, 130, 3–28.
- Malt, B. C., & Smith, E. E. (1984). Correlated properties in natural categories. *Journal of Verbal Learning and Verbal Behavior*, 23, 250–269.
- Mandler, J. M., & Bauer, P. J. (1988). The cradle of categorization: Is the basic level basic? *Cognitive Development*, 3, 247–264.
- Mandler, J. M., Bauer, P. J., & McDonough, L. (1991). Separating the sheep from the goats: Differentiating global categories. *Cognitive Psychology*, 23, 263–298.
- Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, 8, 291–318.
- McClelland, J. L., & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience*, 4, 310–322.
- McRae, K., Cree, G. S., Westmacott, R., & De Sa, V. R. (1999). Further evidence for feature correlations in semantic memory. *Canadian Journal of Experimental Psychology*, 53, 360–373.
- McRae, K., De Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126, 99–130.
- McRae, K., & Jones, M. N. (2013). Semantic memory. In D. Reisberg (Ed.), *The Oxford handbook of cognitive psychology* (pp. 206–219). Oxford, UK: Oxford University Press.
- Medin, D. L., Lynch, E. B., Coley, J. D., & Atran, S. (1997). Categorization and reasoning among tree experts: Do all roads lead to Rome? *Cognitive Psychology*, 32, 49–96.
- Nelson, K. (1977). The syntagmatic-paradigmatic shift revisited: A review of research and theory. *Psychological Bulletin*, 84, 93–116.
- Nguyen, S. P. (2007). Cross-classification and category representation in children's concepts. *Developmental Psychology*, 43, 719–731.
- Nguyen, S. P., & Murphy, G. L. (2003). An apple is more than just a fruit: Cross-classification in children's concepts. *Child Development*, 74, 1783–1806.
- Oakes, L. M., Plumert, J. M., Lansink, J. M., & Merryman, J. D. (1996). Evidence for task-dependent categorization in infancy. *Infant Behavior and Development*, 19, 425–440.
- Pauen, S. (2002). The global-to-basic shift in infants' categorical thinking: First evidence from a longitudinal study. *International Journal of Behavioural Development*, 26, 492–499.
- Peters, R., & Borovsky, A. (2019). Modeling early lexico-semantic network development: Perceptual features matter most. *Journal of Experimental Psychology: General*, 148, 763–782.
- R Core Team. (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. <http://www.R-project.org/>.
- Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. Cambridge, MA: MIT Press.
- Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 7, 573–605.
- Ryan, T. A. (1959). Comments on orthogonal components. *Psychological Bulletin*, 56, 394–396.
- Sizemore, A. E., Karuza, E. A., Giusti, C., & Bassett, D. S. (2018). Knowledge gaps in the early growth of semantic feature networks. *Nature Human Behaviour*, 2, 682–692.
- Smiley, S. S., & Brown, A. L. (1979). Conceptual preference for thematic or taxonomic relations: A nonmonotonic age trend from preschool to old age. *Journal of Experimental Child Psychology*, 28, 249–257.
- Suzuki, R., & Shimodaira, H. (2006). Pvcust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 22, 1540–1542.
- Unger, L., & Fisher, A. V. (2019). Rapid, experience-related changes in the organization of children's semantic knowledge. *Journal of Experimental Child Psychology*, 179, 1–22.
- Unger, L., Fisher, A. V., Nugent, R., Ventura, S. L., & MacLellan, C. J. (2016). Developmental changes in semantic knowledge organization. *Journal of Experimental Child Psychology*, 146, 202–222.
- Vales, C., & Fisher, A. V. (2019). When stronger knowledge slows you down: Semantic relatedness predicts children's co-activation of related items in a visual search paradigm. *Cognitive Science*, 43, e12746.

- Vales, C., States, S. L., & Fisher, A. V. (2020). Experience-driven semantic differentiation: Effects of a naturalistic experience on within- and across-domain differentiation in children. *Child Development*, 91, 733–742.
- Walsh, M., Richardson, K., & Faulkner, D. (1993). Perceptual, thematic, and taxonomic relations in children's mental representations: Responses to triads. *European Journal of Psychology of Education*, 8, 85–102.
- Waxman, S. R., & Namy, L. L. (1997). Challenging the notion of a thematic preference in young children. *Developmental Psychology*, 33, 555–567.